Analogues of a Theorem of Schur on Matrix Transformations*

H. J. Ryser 1971

California Institute of Technology, Pasadena, California 91109

Received

1. Introduction

Let A and B be matrices of sizes m by t and t by n, respectively, with elements in a field F. Let $x_1, \ldots, x_t$ denote t independent indeterminates over F and define

$$X = \mathrm{diag}[x_1, \ldots, x_t]. \tag{1.1}$$

Then

$$AXB = Y \tag{1.2}$$

is a matrix of size m by n such that every element of Y is a linear form in $x_1, \ldots, x_t$ over F. In the present paper we investigate the converse proposition. Thus let

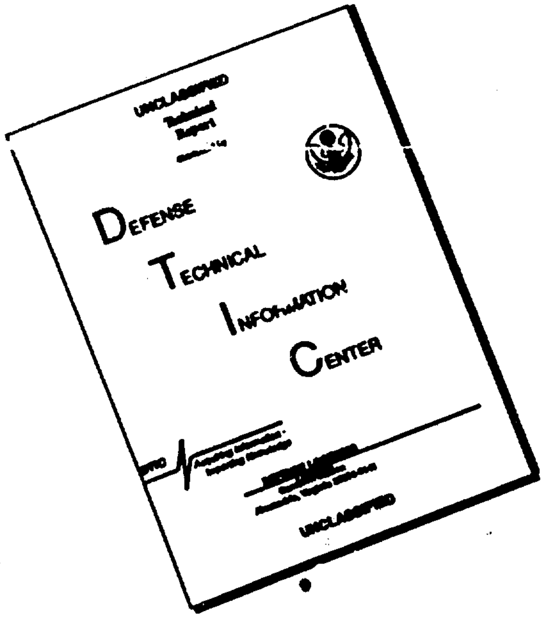
$$Y = Y(x_1, \ldots, x_t) \tag{1.3}$$

be a matrix of size m by n such that every element of Y is a linear form in $x_1, \ldots, x_t$ over F. Then under what conditions are we assured

*This research was supported in part by the Office of Naval Research under Contract N00014-67-A-0094-0010.

15

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Proofs to:
H. J. Ryser
Department of Mathematics
California Institute of Technology
Pasadena, California 91109

of the existence of a factorization of Y of the form (1.2)? Our conditions turn out to be very natural ones and they are easily described in terms of compound matrices. We now state in entirely elementary terms a special case of one of our conclusions.

Theorem 1.1. Let Y be a matrix of order $n \geq 3$ such that every element of Y is a linear form in $x_1, \ldots, x_n$ over F and let

$$X = \operatorname{diag}[x_1, \ldots, x_n]. \tag{1.4}$$

Suppose that

$$\det(Y) = cx_1 \cdots x_n, \tag{1.5}$$

where $c \neq 0$ and $c \in F$, and suppose further that every element of Y^{-1} is a linear form in $x_1^{-1}, \ldots, x_n^{-1}$ over F. Then there exist matrices A and B of order n with elements in F such that

$$AXB = Y. \tag{1.6}$$

Our work has been strongly motivated by the much earlier investigations of Kantor [2], Frobenius [1], and Schur [5]. These authors study a related problem but with X a matrix of size m by n and such that the elements of X are mn independent variables over the complex field. A more recent account of this theory is available in [3].

Finally, we remark that the matrix equation (1.2) is of considerable combinatorial importance in its own right. For example, if A and B are $(0,1)$-matrices, then (1.2) admits of a simple set theoretic interpretation.

The special case

$$AXA^T = Y, \tag{1.7}$$

where A^T is the transpose of A, has been investigated briefly in [4]. But we do not pursue the combinatorial aspects of this subject here.

2. The Main Theorems

Throughout the discussion we let F denote an arbitrary field and we let $x_1, \ldots, x_t$ denote t independent indeterminates over F. We define

$$X = \mathrm{diag}[x_1, \ldots, x_t]. \tag{2.1}$$

We then form all of the products of $x_1, \ldots, x_t$ taken r at a time and we always denote these products written for convenience in the "lexicographic" ordering by

$$y_1, \ldots, y_u \qquad (u = \binom{t}{r}). \tag{2.2}$$

Now let

$$Y = Y(x_1, \ldots, x_t) \tag{2.3}$$

denote a matrix of size m by n such that every element of Y is a linear form in $x_1, \ldots, x_t$ over F. We further assume that

$$1 \leqq r \leqq \min(m,n) \tag{2.4}$$

and we let $C_r(Y)$ denote the rth compound of the matrix Y. Thus $C_r(Y)$ is of size $\binom{m}{r}$ by $\binom{n}{r}$ and the elements of $C_r(Y)$ are the determinants

of the various submatrices of order r of Y displayed within $C_r(Y)$ in the "lexicographic" ordering. We note that the preceding terminology implies

$$C_r(X) = \text{diag}[y_1, \ldots, y_u]. \tag{2.5}$$

We are now prepared to state one of our main conclusions.

Theorem 2.1. <u>Let</u> Y <u>denote a matrix of size</u> m <u>by</u> n <u>such that every element of</u> Y <u>is a linear form in</u> $x_1, \ldots, x_t$ <u>over</u> F <u>and let</u> $y_1, \ldots, y_u$ <u>denote the products of</u> $x_1, \ldots, x_t$ <u>taken</u> r <u>at a time</u>. <u>We assume that</u>

$$2 \leqq r \leqq \text{rank }(Y) - 2 \tag{2.6}$$

<u>and that every element of</u> $C_r(Y)$ <u>is a linear form in</u> $y_1, \ldots, y_u$ <u>over</u> F. <u>Then there exist matrices</u> A <u>and</u> B <u>of sizes</u> m <u>by</u> t <u>and</u> t <u>by</u> n, <u>respectively, with elements in</u> F <u>such that</u>

$$AXB = Y. \tag{2.7}$$

We begin with a simple lemma concerning the matrix Y of (2.3).

Lemma 2.2. <u>Let</u>

$$Y_i = Y(0, \ldots, 0, x_i, 0, \ldots, 0) \tag{2.8}$$

<u>and suppose that</u>

$$\text{rank }(Y_i) \leqq 1 \quad (i = 1, \ldots, t). \tag{2.9}$$

<u>Then there exist matrices</u> A <u>and</u> B <u>of sizes</u> m <u>by</u> t <u>and</u> t <u>by</u> n, <u>respectively,</u>

<u>with elements in</u> F <u>such that</u>

$$AXB = Y. \tag{2.10}$$

<u>Proof</u>. The assertion rank $(Y_i) \leqq 1$ implies that we may write

$$Y_i = \alpha_i x_i \beta_i, \tag{2.11}$$

where

$$\alpha_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix}, \quad \beta_i = (b_{i1}, \ldots, b_{in}) \tag{2.12}$$

are vectors with components in F. Here if rank $(Y_i) = 1$ we have $\alpha_i \neq 0$ and $\beta_i \neq 0$. But if rank $(Y_i) = 0$ we have $\alpha_i = 0$ and β_i arbitrary or $\beta_i = 0$ and α_i arbitrary. Thus

$$\begin{aligned} Y = Y_1 + \cdots + Y_t &= \alpha_1 x_1 \beta_1 + \cdots + \alpha_t x_t \beta_t \\ &= [\alpha_1, \ldots, \alpha_t] \, X \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_t \end{bmatrix}, \end{aligned} \tag{2.13}$$

and our conclusion follows.

Notice further that if

$$\text{rank } (Y_i) = 1 \quad (i = 1, \ldots, t) \tag{2.14}$$

and if

$$A'XB' = Y, \tag{2.15}$$

then there exists a nonsingular diagonal matrix D with elements in F such that

$$A' = AD^{-1}, \quad B' = DB. \tag{2.16}$$

It is now clear that the following lemma is actually a reformulation of Theorem 2.1.

Lemma 2.3. <u>The matrix</u> Y <u>of</u> Theorem 2.1 <u>satisfies</u>

$$\text{rank } (Y_i) \leqq 1 \quad (i = 1, \ldots, t). \tag{2.17}$$

<u>Proof</u>. We remark at the outset that the lemma is elementary for $r = 2$. In this case rank $(Y_i) \leqq 1$ because otherwise we contradict the assumption that every element of $C_2(Y)$ is a linear form in $y_1, \ldots, y_u$ over F.

Hence we take $r \geqq 3$. Let us suppose that

$$\text{rank } (Y_i) = p > 1 \tag{2.18}$$

for some $i = 1, \ldots, t$. Then there exist nonsingular matrices P and Q of orders m and n, respectively, with elements in F such that

$$PY_iQ = x_iI \oplus 0. \tag{2.19}$$

In (2.19) the matrix I is the identity matrix of order p, 0 is a zero matrix, and the sum is direct. The elements of the matrix

$$PYQ = Z \tag{2.20}$$

are linear forms in $x_1, \ldots, x_t$ over F. It follows from (2.13) and

(2.19) that the structure of Z is such that the indeterminate x_i appears in positions (1,1), ..., (p,p), and in no other positions in Z. The familiar multiplicative property of the compound matrix implies

$$C_r(P)C_r(Y)C_r(Q) = C_r(Z), \tag{2.21}$$

and by our assumption on $C_r(Y)$ we may conclude that each of the elements of $C_r(Z)$ is also a linear form in $y_1, \ldots, y_u$ over F.

We designate by F_i the quotient field of the polynomial ring

$$F[x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_t]. \tag{2.22}$$

In this notation the elements of Z and $C_r(Z)$ are scalars or polynomials in x_i of degree 1 over F_i. In what follows we apply certain elementary row and column operations to Z with respect to the field F_i. This means that we determine certain nonsingular matrices P' and Q' of orders m and n, respectively, with elements in F_i such that

$$P'ZQ' = Z'. \tag{2.23}$$

Then once again we have

$$C_r(P')C_r(Z)C_r(Q') = C_r(Z'). \tag{2.24}$$

Thus we see that the elements of Z' and $C_r(Z')$ are scalars or polynomials in x_i of degree 1 over F_i.

We now write Z in the form

$$Z = \begin{bmatrix} W & * \\ * & * \end{bmatrix}, \tag{2.25}$$

where W is of order p. We note that det(W) is a polynomial in x_i of degree $p > 1$ over F_i. Let the submatrix of Z in the lower right corner of Z of size m - p by n - p be of rank ρ. Then we may apply elementary row and column operations with respect to F_i to the last m - p rows and the last n - p columns of Z and replace Z by

$$Z' = \begin{bmatrix} W & * & * \\ * & I & 0 \\ W' & 0 & 0 \end{bmatrix} . \tag{2.26}$$

In (2.26) the matrix I is the identity matrix of order ρ and the 0's denote zero matrices. We assert that

$$p + \rho \leq r - 1 \tag{2.27}$$

because $p + \rho \geq r$ contradicts the fact that all of the elements of $C_r(Z')$ are scalars or polynomials in x_i of degree 1 over F_i. Let the submatrix W′ of Z′ be of rank ρ'. We have rank (Z′) = rank (Y) and hence we may conclude that

$$p + \rho + \rho' \geqq \text{rank } (Y). \tag{2.28}$$

It now follows from (2.6), (2.27), and (2.28) that

$$\rho' \geqq 3. \tag{2.29}$$

We permute the last m - (p + ρ) rows and the first p columns of Z′ so that the submatrix of order 2 in the lower left corner of W′ has a nonzero determinant. We then further permute the first p rows of Z′

so that the p polynomials in x_i of degree 1 over F_i again occupy the main diagonal positions of W. By elementary row operations with respect to F_i we may replace the matrix of order 2 in the lower left corner of W' by the identity matrix. We then apply further elementary row operations with respect to F_i and make all elements in columns 1 and 2 of Z' equal to 0, apart from the elements in the (1,1), (2,2), (m-1,1), (m,2) positions, and these elements are equal to x_i, x_i, 1, 1, respectively.

We delete rows 1, 2, m-1, m and columns 1, 2 from Z' and call the resulting submatrix $\widetilde{Z}$. Then we have

$$Z' = \left[\begin{array}{cc|c} x_i & 0 & \\ 0 & x_i & * \\ \hline 0 & 0 & \\ \vdots & \vdots & \widetilde{Z} \\ 0 & 0 & \\ \hline 1 & 0 & \\ 0 & 1 & * \end{array}\right]. \tag{2.30}$$

The matrix $\widetilde{Z}$ is of size m - 4 by n - 2. Let $\widetilde{Z}$ be of rank $\widetilde{\rho}$. We have rank (Z') = rank (Y) and hence

$$\widetilde{\rho} + 4 \geq \text{rank } (Y). \tag{2.31}$$

We assert that

$$c_{r-2}(\widetilde{Z}) \neq 0. \tag{2.32}$$

Suppose on the contrary that $C_{r-2}(\tilde{Z}) = 0$. Then

$$\tilde{\rho} \leqq r - 3. \tag{2.33}$$

But then by (2.6), (2.31), and (2.33) we have

$$\text{rank } (Y) \leqq \tilde{\rho} + 4 \leqq r + 1 \leqq \text{rank } (Y) - 1, \tag{2.34}$$

and this is a contradiction. Hence $C_{r-2}(\tilde{Z}) \neq 0$. This means that $\tilde{Z}$ has a submatrix of order $r-2$ with a nonzero determinant. But this submatrix of $\tilde{Z}$ in conjunction with the first two rows and columns of Z' yields a submatrix of Z' of order r whose determinant is a polynomial in x_i of degree 2 or higher over F_i. This contradicts the fact that the elements of $C_r(Z')$ are scalars or polynomials in x_i of degree 1 over F_i. Hence we have

$$\text{rank } (Y_i) \leqq 1 \quad (i = 1, \ldots, t). \tag{2.35}$$

This proves Lemma 2.3 and Theorem 2.1.

The range of r in the preceding theorem cannot in general be extended to $r = \text{rank } (Y) - 1$. We define

$$Y = \text{diag}[x_1, \ldots, x_n] + \left[\begin{array}{cc|c} \multicolumn{2}{c|}{0} & 0 \\ \hline x_{n+1} & 0 & 0 \\ 0 & x_{n+1} & \end{array}\right], \tag{2.36}$$

where the 0's denote zero matrices. Then we have $t = n + 1$ and if

$n \geqq 4$ we have

$$Y^{-1} = \operatorname{diag}\left[\frac{1}{x_1}, \ldots, \frac{1}{x_n}\right] + \left[\begin{array}{cc|c} \multicolumn{2}{c|}{0} & 0 \\ \hline \frac{-x_{n+1}}{x_1x_{n-1}} & 0 & \\ & & 0 \\ 0 & \frac{-x_{n+1}}{x_2x_n} & \end{array}\right]. \quad (2.37)$$

Hence for $r = n - 1$ we see that every element of $C_r(Y)$ is a linear form in $y_1, \ldots, y_u$ over F. But clearly rank $(Y_{n+1}) = 2$.

The preceding theorem, however, is valid for r = rank $(Y) - 1$ under the added assumption t = rank (Y). This theorem is actually a generalization of Theorem 1.1 described in Section 1.

Theorem 2.4. Let Y denote a matrix of size m by n such that every element of Y is a linear form in $x_1, \ldots, x_t$ over F and let $y_1, \ldots, y_u$ denote the products of $x_1, \ldots, x_t$ taken r at a time. We assume that

$$2 \leqq r = \text{rank } (Y) - 1, \quad (2.38)$$

$$t = \text{rank } (Y), \quad (2.39)$$

and that every element of $C_r(Y)$ is a linear form in $y_1, \ldots, y_u$ over F. Then there exist matrices A and B of sizes m by t and t by n, respectively, with elements in F such that

$$AXB = Y. \quad (2.40)$$

Lemma 2.5. *Let* Y *be a nonsingular matrix of order* $t \geq 3$ *such that every element of* Y *is a linear form in* $x_1, \ldots, x_t$ *over* F. *Let* $r = t - 1$ *and suppose that every element of* $C_r(Y)$ *is a linear form in* $y_1, \ldots, y_u$ *over* F. *Then*

$$\det(Y) = cx_1 \cdots x_t, \tag{2.41}$$

where $c \neq 0$ *and* $c \in F$.

Proof. Let

$$\text{rank } (Y_i) = p. \tag{2.42}$$

We apply the same elementary row and column operations as in Lemma 2.3. Thus we know that there exist nonsingular matrices P and Q of order t with elements in F such that

$$PYQ = Z. \tag{2.43}$$

The elements of Z are linear forms in $x_1, \ldots, x_t$ over F. But the structure of Z is such that x_i appears in positions $(1,1), \ldots, (p,p)$, and in no other positions in Z. We know that every element of $C_r(Z)$ is a linear form in $y_1, \ldots, y_u$ over F. Hence $t \geq 3$ implies that we cannot have x_i in the (t,t) position of Z. Thus x_i does not occur in the last column of Z. An evaluation of $\det(Z)$ by this column implies that no term of $\det(Z)$ contains x_i to a power higher than the first. Thus no term of $\det(Y)$ contains x_i to a power higher than the first, and this is valid for each $i = 1, \ldots, t$. Hence by the structure of Y we conclude that $\det(Y)$ is a nonzero scalar multiple of $x_1 \cdots x_t$.

The following lemma completes the proof of Theorem 2.4.

Lemma 2.6. <u>The matrix</u> Y <u>of</u> Theorem 2.4 <u>satisfies</u>

$$\text{rank } (Y_i) \leqq 1 \quad (i = 1, \ldots, t). \tag{2.44}$$

<u>Proof</u>. We assume that

$$\text{rank } (Y_i) = p > 1 \tag{2.45}$$

for some $i = 1, \ldots, t$. Once again there exist nonsingular matrices P and Q of orders m and n, respectively, with elements in F such that

$$PYQ = Z. \tag{2.46}$$

The elements of Z are linear forms in $x_1, \ldots, x_t$ over F. But the structure of Z is such that the indeterminate x_i appears in positions $(1,1), \ldots, (p,p)$, and in no other positions in Z. Furthermore, every element of $C_r(Z)$ is a linear form in $y_1, \ldots, y_u$ over F.

The submatrix W of order p in the upper left corner of Z is nonsingular because its determinant is a polynomial in x_i of degree p over F_i. We have

$$t = \text{rank } (Y) = \text{rank } (Z) \geqq 3 \tag{2.47}$$

and hence Z contains a nonsingular submatrix Z' of order t with W in its upper left corner. We now write

$$Z'Z'^{-1} = I. \tag{2.48}$$

The elements of Z' are of the form $ax_i + b$, where $a,b \in F_i$. Moreover,

the polynomials in x_i of degree 1 over F_i appear in positions $(1,1), \ldots, (p,p)$, and in no other positions in Z'. Every element of $C_r(Z')$ is a linear form in $y_1, \ldots, y_u$ over F. Hence by Lemma 2.5 every element of Z'^{-1} is of the form $cx_i^{-1} + d$, where $c,d \in F_i$. We now multiply row 1 of Z' by column j of Z'^{-1}. This product is 0 or 1. Hence the element in the $(1,j)$ position of Z'^{-1} is of the form cx_i^{-1}, where $c \in F_i$. Similarly, each of the elements in the first p rows of Z'^{-1} is of this form. Hence

$$\det(Z'^{-1}) = x_i^{-p} f(x_i^{-1}), \tag{2.49}$$

where $f(x_i^{-1})$ is a nonzero polynomial in x_i^{-1} over F_i. But by Lemma 2.5 we have

$$\det(Z'^{-1}) = ex_i^{-1}, \tag{2.50}$$

where $e \neq 0$ and $e \in F_i$. This contradicts $p > 1$. Hence $p = 1$ and the lemma is established.

References

1. G. Frobenius, Über die Darstellung der endlichen Gruppen durch lineare Substitutionen, _Sitzungsberichte Berliner Akademie_ (1897), 994-1015.

2. S. Kantor, Theorie der Äquivalenz von linearen ∞^λ-Scharen bilinearer Formen, _Sitzungsberichte Münchener Akademie_ (1897), 367-381.

3. M. Marcus and F. May, On a theorem of I. Schur concerning matrix transformations, _Archiv. Math._ 11 (1960), 401-404.

4. H. J. Ryser, A fundamental matrix equation for finite sets, (submitted).

5. I. Schur, Einige Bemerkungen zur Determinantentheorie, _Sitzungsbericnte Berliner Akademie_ (1925), 454-463.